7Б

УДК 519.86:519.857.3:519.217

І.В. Баклан, Г.А. Степанкова
Національна академія управління, м. Київ, Україна
iii2@online.ua, astepankova@mail.ru

Імовірнісні моделі для аналізу та прогнозування часових рядів

Досліджені основні результати використання апарату прихованих марківських моделей для аналізу часових рядів та пов'язані з ними діаграми впливу. Розглянуті особливості використання апарату марківських ланцюжків для аналізу часових рядів та навчання марківських ланцюжків.

Вступ

В українській мові є чимало слів, які позначають спробу зазирнути у майбутнє: передчуття, інтуїція, прогнозування, ознака, передбачення, пророкування, пророцтво, віщування, гадання та таке інше. До математичних методів має відношення, мабуть, тільки прогнозування та передбачення.

Людину завжди цікавило майбутнє. Цей інтерес зрозумілий та не позбавлений здорового глузду – точна інформація про розмір ринкового попиту, зміни цін, курсів валют через тиждень, місяць, рік надають володарю знань можливість впливати на ситуацію сьогодні для отримання прибутку, дивідендів та преференцій у майбутньому. Оскільки економічні умови здійснення бізнесу змінюються з часом, менеджерам особливо вимагається «тримати руку на пульсі» цих змін для результативного виконання ділових операцій.

Одним із заходів, яким менеджери можуть скористатися при оцінці ефективності майбутніх керівних рішень, є прогнозування. На цей час розроблено багато методів прогнозування, у яких загальна мета – завбачити з тим чи іншим ступенем надійності майбутні події та врахувати цей прогноз при плануванні управлінських рішень.

Формально використовується два підходи до прогнозування – якісний та кількісний. Методи якісного прогнозування важливі, коли дані за минулі проміжки часу недоступні або ненадійні. Наприклад, при прогнозуванні обсягу продажу цілком нового товару, який не існував ще на ринку. Усі якісні методи (наприклад, метод експертної оцінки, дельфійський метод) занадто суб'єктивні і тому не досить достовірні. Кількісні методи базуються на використанні інформації за минулі періоди часу; при дослідженні тенденцій процесу в минулому можна з'ясувати основні взаємозалежності між величинами та видати більш надійний прогноз на майбутнє.

Усі методи кількісного прогнозування також можливо розподілити на два типи: причинні методи та методи, які базуються на аналізі часових рядів. Перші з них (що звуть часто методами моделювання процесів) мають у складі визначення значимих чинників та функціональної залежності відгуку від цих чинників із застосуванням множинного регресійного аналізу або економетричного моделювання (факторний аналіз). Прогноз за часовими рядами, в свою чергу, передбачає визначення прогнозного значення змінної виключно на підставі минулих та поточних значень цієї ж змінної. В найпростішому випадку прогнозування може бути результативно виконано за допомогою звичайного спостереження за явищем з цікавлячої нас сторони.

Прогнозування на основі часових рядів – один із самих популярних підходів до прогнозування розвитку економічних процесів, об'ємів торгових операцій, об'ємів виробництва та накопичення продукції на складах, оцінювання альтернативних економічних стратегій, формування бюджетів підприємств та держави, прогнозування та менеджмент економічних і фінансових ризиків та інше.

Спочатку розглянемо основні визначення та результати щодо використання апарату прихованих марківських моделей для аналізу часових рядів [1], [2]. Буде розглянутий апарат марківських ланцюжків. Розглянемо особливості використання апарату марківських ланцюжків для аналізу часових рядів. Дамо класифікацію марківських моделей. Також розглянемо приховані марківські моделі (ПММ) та пов'язані з ними діаграми впливу.

Марківські ланцюжки

Розглянемо алфавіт $S = \{s_1, s_2, \ldots, s_J\}$ та послідовність випадкових величин $X_0, X_1, \ldots, X_n, \ldots$, які приймають значення на цьому алфавіті S. Стан s_j в послідовності, яку називають множиною або простором станів. Для зручності надалі стани будемо позначати їх індексами і тоді $S = \{1, 2, \ldots, J\}$.

Послідовність випадкових величин $\{X_n\}_{n=0}^{\infty}$ називається *марківським ланцюжком* (МЛ), якщо для всіх $n \geq 1$ та $j_1, j_2, \ldots, j_n \in S$ має місце:

$$P\left(X_n = j_n \middle| X_0 = j_0, \ldots, X_{n-1} = j_{n-1}\right) = P\left(X_n = j_n \middle| X_{n-1} = j_{n-1}\right). \tag{1}$$

Зміст МЛ полягає в тому, що якщо $X_n = j_n$ – майбутній випадок, то умовна ймовірність цього випадку, яка задається історичною послідовністю $X_0 = j_0$, $X_1 = j_1, \ldots, X_{n-1} = j_{n-1}$, залежить тільки від свого безпосереднього попередника $X_{n-1} = j_{n-1}$ та не залежить від $X_0 = j_0, X_1 = j_1, \ldots, X_{n-2} = j_{n-2}$.

Рівняння, наведене в (1), відоме як м*арківська властивість*.

Нехай $\{X_n\}_{n=0}^{\infty}$ – МЛ. Якщо $X_n = j$, то ми кажемо, що ланцюжок знаходиться у стані j на момент часу n або ланцюжок у момент часу n перейшов до стану j. Тоді умовні ймовірності

$$p_{i|j} = P\left(X_n = j \middle| X_{n-1} = i\right),\ n \geq 1,\ i, j \in S. \tag{2}$$

Виникає ряд запитань. В яких випадках МЛ має унікальний інваріантний розподіл? Чи існує збіжність до інваріантного розподілу? У загальному випадку на ці запитання досить важко відповісти. Але для кінцевих просторів станів відповіді можуть бути отримані. Надалі нас будуть цікавити марківські ланцюжки цього типу.

Про МЛ $\{X_n\}_{n=0}^{\infty}$ говорять, що він *ергодичний*, якщо існує розподіл ймовірностей $a = (a_1, \ldots, a_J)$, заданий на фазовому просторі S, такий, що

$$\lim_{n \to \infty} \pi(n) = a \tag{3}$$

для будь-якого початкового розподілу $\pi(0)$.

Безсумнівно, що ліміт-розподіл a повинен бути унікальним. Для ергодичного МЛ ліміт-розподіл та інваріантний розподіл ідентичні.

Наведемо *умови достатності ергодичності*:

Припустимо, що $\lim\limits_{n \to \infty} p_{i|j}(n) = a_j$ для усіх j та незалежно від i. Візьмемо $n = 1$ та, припускаючи, що в рівнянні Чапмена-Колмогорова $m \to \infty$, ми отримаємо

$$a_j = \lim_{m \to \infty} p_{i|j}(m+1) = \sum_{k=1}^{J} \lim_{m \to \infty} p_{k|j}(m) \cdot p_{k|j}(1) =$$

$$= \sum_{k=1}^{J} a_k \cdot p_{k|j}(1) = \sum_{k=1}^{J} a_k \cdot p_{k|j} ,$$

оскільки $p_{k|j}(1) = p_{k|j}$ для першого кроку. Інакше кажучи,

$$a_j = \sum_{k=1}^{J} a_k \cdot p_{k|j} ,\ j = 1, \ldots, J . \quad (4)$$

Останній вираз може бути записаний як $a = aP$. Таким чином, ми бачимо, що лімітні властивості матриць переходів n кроків також пов'язані з питанням лімітації розподілів.

Навчання марківських ланцюжків

Тепер ми будемо застосовувати кроки байєсова моделювання до тренування послідовності, яка використовує родину марківських моделей. Марківська модель повністю визначається матрицею переходів та початковим розподілом ймовірностей. Імовірнісне навчання передбачає:

1. Марківський імовірнісний розподіл визначає ймовірність будь-якої послідовності, обумовленої матрицею переходів та початковим розподілом.

2. Передування, яке виражає невизначеність про матрицю переходів та початковий розподіл.

Якщо (1) об'єднано відомим чином з процесом тренування послідовності, ми отримуємо функцію правдоподібності, що стосується родини марківських моделей. Функція правдоподібності комбінована з (2) через байєсівські правила породжує апостеріорний розподіл для параметрів родини марківських моделей. Використовуючи (1) та (2), ми також обчислюємо прогнозований розподіл та отримуємо порівняльну модель за допомогою байєсівських чинників.

Розглянемо *максимальну достовірність* для марківських ланцюжків.

Нехай $\underline{\theta}$ – матриця ймовірностей переходів:

$$\underline{\theta} = \begin{pmatrix} \theta_{1|1} & \theta_{1|2} & \ldots & \theta_{1|J} \\ \theta_{2|1} & \theta_{2|2} & \ldots & \theta_{2|J} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{J|1} & \theta_{J|2} & \ldots & \theta_{J|J} \end{pmatrix} . \quad (5)$$

Нас турбує оцінка моделі $\underline{\theta}$ в родині імовірнісних моделей $p(\mathbf{x}|\underline{\theta})$ для навчання послідовності $\mathbf{x}$ з $n+1$ станів в S^{n+1}

$$\mathbf{x} = (j_0 j_1 \ldots j_n) \in S^{n+1} .$$

Як було показано раніше, для МЛ

$$p(\mathbf{x}|\underline{\theta}) = P(X_0 = j_0, X_1 = j_1, X_n = j_n|\theta) = \pi_{j_0}(0) \prod \theta_{j_{l-1}|j_l} . \quad (6)$$

Тоді під *моделлю родини* будемо розуміти обумовлену на $\pi_{j_0}(0)$ та $\Theta = \underline{\theta}$, послідовність станів в $\mathbf{x}$, що є виходом Марківського ланцюжку $\{X_n\}_{n \geq 0}$ із стаціонарними ймовірностями переходу $\underline{\theta}$.

Тут ми маємо щонайбільше $J^2 - J$ параметрів переходу та $J-1$ початкових ймовірностей для оцінки використання даних $\mathbf{x}$. Ми робимо апроксимацію з тим, щоб знехтувати початковим розподілом як частиною проблеми оцінювання.

Як функцію $\underline{\theta}$ для фіксованого $\mathbf{x}$ забезпечення апроксимації або умовної функції правдоподібності візьмемо

$$L(\underline{\theta}) = \prod_{l=1}^{n} \theta_{j_{l-1}|j_l} . \tag{7}$$

Відповідна логарифмічна функція правдоподібності:

$$\ell(\underline{\theta}) = \sum_{l=1}^{n} \ln \theta_{j_{l-1}|j_l} . \tag{8}$$

Тепер розглянемо *максимальну достовірність матриці переходів*.

Введемо запис для кількості часу, який вимагається для переходу від i до j в $\mathbf{x} = (j_0 j_1 \dots j_n)$. Маємо

$$n_{i|j} = \text{кількість таких } l \text{, що } 1 \le l \le n,\ j_{l-1} = i,\ j_l = j . \tag{9}$$

Використовуючи підрахунок частоти $n_{i|j}$, ми можемо записати функцію достовірності таким чином:

$$L(\underline{\theta}) = \prod_{i=1}^{J} \prod_{j=1}^{J} \theta_{i|j}^{n_{i|j}} . \tag{10}$$

Без сумніву, всі $n_{i|j}$ – достатні статистики для цієї моделі родини. Логарифмічна достовірність з (8) буде тепер мати вигляд:

$$\ell(\underline{\theta}) = \sum_{i=1}^{J} \sum_{j=1}^{J} n_{i|j} \ln \theta_{i|j} . \tag{11}$$

Нехай також

$$n_i = \text{кількість таких } l \text{, що } 1 \le l \le n,\ j_{l-1} = i,\ j_l = j , \tag{12}$$

так, що n_i дорівнює кількості часу, за який досягається стан i в послідовності $\mathbf{x}$, в тому числі й перехід за останній проміжок часу.

Розглянемо *усереднену модель* та *апостеріорний розподіл* для рядків матриці переходу.

Припустимо, що невизначеність ряду $\underline{\theta}$ в (5)

$$\theta_i = \left(\theta_{i|1}, \dots, \theta_{i|J}\right)$$

моделюється незалежними спадковими величинами, що мають відповідну щільність Діріхле для $i = 1, \dots, J$.

Нехай $S_L \subset R^L$ буде симплекс

$$S_L = \left\{ (\theta_1, \dots, \theta_L) \middle| \theta_i \ge 0, \quad i = 1, \dots, L, \quad \sum_{i=1}^{L} \theta_i = 1 \right\}.$$

Нехай для $\alpha_i > 0$

$$\phi(\theta_1, \dots, \theta_L) = \begin{cases} \dfrac{\prod_{i=1}^{L} \theta_i^{\alpha_i - 1}}{Z}, & \textit{якщо} \quad \theta_1, \dots, \theta_L \in S_L, \\ 0 & \textit{у протилежному випадку}. \end{cases}$$

Тут

$$\frac{1}{Z} = \frac{\Gamma\left(\sum_{i=1}^{L} \alpha_i\right)}{\prod_{i=1}^{L} \Gamma(\alpha_i)} .$$

В цьому випадку щільність $\phi(\theta_1,\ldots,\theta_L)$ називають щільністю Діріхле. За звичай її позначають

$$Dir(\alpha_1,\ldots,\alpha_L).$$

Якщо $\alpha_1 = \alpha_2 = \ldots = \alpha_L = k$, то будемо говорити про симетричну щільність Діріхле. Щільність Діріхле має властивість

$$\int_{S_L} \phi(\theta_1,\ldots,\theta_L) d\theta_1 \ldots d\theta_L = 1,$$

доведення якої було наведене у Вілкса [3].

В нашому випадку маємо:

$$Dir\left(\theta_i;\alpha_1\cdot q_{i|1},\ldots,\alpha_1\cdot q_{i|J}\right) = \frac{\Gamma(\alpha_i)}{\prod_{j=1}^{J}\Gamma\left(\alpha_i q_{i|j}\right)}\cdot\prod_{j=1}^{J}\theta_{i|j}^{\alpha_i q_{i|j}-1}, \tag{13}$$

де

$$\alpha_i > 0,\; q_{i|j} > 0,\; \sum_{j=1}^{J} q_{i|j} = 1. \tag{14}$$

Тоді ми будемо використовувати у якості щільності спільного апріорного розподілу щільність багатомірного розподілу Діріхле:

$$\prod_{i=1}^{J} Dir\left(\theta_i;\alpha_1\cdot q_{i|1},\ldots,\alpha_1\cdot q_{i|J}\right). \tag{15}$$

Таким чином, щільність апостеріорного розподілу дорівнює:

$$p(\underline{\theta}|x) = \frac{\prod_{i=1}^{J}\frac{\Gamma(\alpha_i)}{\prod_{j=1}^{J}\Gamma\left(\alpha_i q_{i|j}\right)}\cdot\prod_{j=1}^{J}\theta_{i|j}^{\alpha_i q_{i|j}-1}}{p(x)}, \tag{16}$$

де $p(x)$ є нормалізація, яка переводить $p(\underline{\theta}|x)$ стохастичну щільність в $\underline{\theta}$.

Тепер розглянемо *прогнозну ймовірність*.

Можно запитати, що таке наша ймовірність

$$P\left(X_{n+1} = j | X_n = i;\mathbf{x}\right),$$

де вказано, що ймовірність базується на підготовленій послідовності $\mathbf{x} = \left(j_0 j_1 \ldots j_n\right) \in S^{n+1}$. Враховуючи наш підхід до марківського моделювання $\mathbf{x}$ в попередньому розділі, однією з відповідей могла би бути

$$\hat{P}_{ML}\left(X_{n+1} = j | X_n = i;\mathbf{x}\right) = \hat{\theta}_{i|j},$$

вбудовуючи оцінку максимальної достовірності ймовірності переходу.

З точки зору спостережності ми повинні розглядати тільки унікальну послідовність $\mathbf{x}$ для $P\left(X_{n+1} = j|\mathbf{x}\right)$ та забезпечувати відповідь як

$$\hat{P}_{ML}\left(X_{n+1} = j|\mathbf{x}\right) = \hat{\theta}_{j_n|j},$$

оскільки j_n – остання перемінна у послідовності.

Існують інші шляхи, як відповісти на поставлені питання нашого дослідження. Використовуючи послідовність $\mathbf{x}$, можемо взяти апостеріорну щільність для $\underline{\theta}$ й тоді забезпечити нову матрицю переходів за допомогою моделі усереднення

$$P^*\left(X_{n+1} = j | X_n = i;\mathbf{x}\right) = \int \theta_{i|j}\, p\left(\underline{\theta}|\mathbf{x}\right) d\underline{\theta}. \tag{17}$$

Застосовуючи (15), отримуємо

$$P^*\left(X_{n+1}=r\middle|X_n=s;\mathbf{x}\right)=\int\theta_{r|s}\,p\left(\underline{\theta}\middle|\mathbf{x}\right)d\underline{\theta}=\frac{\prod_{i=1}^{J}\frac{\Gamma(\alpha_i)}{\prod_{j=1}^{J}\Gamma\left(\alpha_i q_{i|j}\right)}\cdot I_{i|j}(r,s)}{p(\mathbf{x})},$$

де

$$I_{i|j}(r,s)=\int\prod_{j=1}^{J}\theta_{r|s}\theta_{i|j}^{n_{i|j}a_i q_{i|j}-1}d\theta_i\ .$$

Використовуючи широковідому формулу для оцінки варіаційних інтегралів Діріхле, ми отримаємо замість (17) наступний вираз

$$P^*\left(X_{n+1}=r\middle|X_n=s;\mathbf{x}\right)=\frac{n_{s|r}+\alpha_s q_{s|r}}{n_s+\alpha_s}. \tag{18}$$

Параметри α_s та $q_{s|r}$, як й раніше, відіграють роль псевдолічильника спостережень та упорядковувача.

Приховані марківські моделі

Приховані марківські моделі (ПММ) (раніше відомі під назвами – стохастична функція марківського ланцюжку або марківське джерело) – випадковий процес, породжений двома взаємозалежними стохастичними механізмами. Це – марківські ланцюжки, що лежать в основі, із скінченою кількістю станів та множиною випадкових функцій, кожна з яких асоційована з відповідним станом. При дискретних моментах часу процес знаходиться в деякому стані та спостереження генерується випадковою функцією, яка відповідає чинному стану. Базовий марківський ланцюжок змінює свій стан відповідно до його матриці переходів.

Спостерігати можливо тільки результат випадкових функцій, прив’язаних до кожного стану, та неможливо прямо спостерігати стани базового марківського ланцюжку. Таким чином, марківський ланцюжок є фактично прихованим, що й дало назву цій родині моделей – прихована марківська модель.

У принципі результатом від станів прихованого марківського ланцюжку може бути багатомірний випадковий процес, який має деяку безперервну функцію спільного розподілу. У цій статті ми обмежимося розглядом процесів із дискретним скінченим алфавітом.

ПММ може розглядатися як родина моделей для послідовності спостережень з послідовності спостережень $O=\{o_1,o_2,\ldots,o_K\}$. Модель базується на ідеї прихованої послідовності переходів марківських станів. Дамо більш формальне визначення ПММ. *Прихований марківський ланцюжок* – це марківський ланцюжок $\{X_n\}_{n=0}^{\infty}$, який приймає значення на скінченому фазовому просторі $S=\{1,2,\ldots,J\}$ з J станів. Умовні ймовірності

$$a_{i|j}=P\left(X_n=j\middle|X_{n-1}=i\right),\ n\geq 1,\ i,j\in S. \tag{19}$$

Матриця переходів будується як

$$A=\left(a_{i|j}\right)_{i=1,\,j=1}^{J,J} \tag{20}$$

із добре відомими нам обмеженнями:

$$a_{i|j}\geq 0,\ \sum_{j=1}^{J}a_{i|j}=1.$$

В момент часу $n = 0$ стан X_0 визначається розподілом ймовірностей $\pi_j(0) = P(X_0 = j)$ з $\pi(0) = (\pi_1(0), \ldots, \pi_J(0))$.

Помітний випадковий процес – це випадковий процес $\{Y_n\}_{n=0}^{\infty}$ із скінченим фазовим простором $O = \{o_1, o_2, \ldots, o_K\}$, де $K \neq J$. Процеси $\{X_n\}_{n=0}^{\infty}$ та $\{Y_n\}_{n=0}^{\infty}$ для довільних фіксованих n пов'язані розподілом умовних ймовірностей

$$b_j(k) = P(Y_n = o_k | X_n = j).$$

Встановимо, що

$$B = \{b_j(k)\}_{j=1,i=1}^{J,K}.$$

Будемо називати її матрицею ймовірностей розповсюдження. Це інша стохастична матриця в тому сенсі, що

$$b_j(k) \geq 0, \ \sum_{k=1}^{K} b_j(k) = 1.$$

Для довільної послідовності станів $j_0 j_1 \ldots j_n$ ймовірність послідовності $o_0 o_1 \ldots o_n$ визначається як:

$$P(Y_0 = o_0, \ldots, Y_n = o_n | X_0 = j_0, \ldots, X_n = j_n, B) = \prod_{l=0}^{n} b_{jl}(l). \tag{21}$$

Інакше кажучи, виділені спостереження умовно незалежні від заданої послідовності станів.

Процес $\{Y_n\}_{n=0}^{\infty}$ може розглядатися як функція від марківського ланцюжку, та в загальному випадку не є марківським ланцюжком.

До набору переваг цих припущень можна записати спільну ймовірність $o_0 o_1 \ldots o_n$ та $j_0 j_1 \ldots j_n$ як

$$\begin{aligned} &P(Y_0 = o_0, \ldots, Y_n = o_n, X_0 = j_0, \ldots, X_n = j_n; A, B, \pi(0)) = \\ &= P(Y_0, \ldots, Y_n | X_0, \ldots, X_n, B) \cdot P(X_0, \ldots, X_n, A, \pi(0)) = \\ &= \pi_{j_0}(0) \cdot \prod_{l=0}^{n} b_{jl}(l) \cdot \prod_{l=1}^{n} a_{j_{l-1}|j_l}. \end{aligned}$$

Перебудовуючи останній вираз, отримаємо:

$$\begin{aligned} &P(Y_0 = o_0, \ldots, Y_n = o_n, X_0 = j_0, \ldots, X_n = j_n; A, B, \pi(0)) = \\ &= \pi_{j_0}(0) \cdot b_{j_0}(0) \cdot \prod_{l=1}^{n} a_{j_{l-1}|j_l} b_{jl}(l). \end{aligned}$$

Надалі ми отримуємо спільний розподіл ймовірностей $o_0 o_1 \ldots o_n$ як неістотний розподіл, підсумовуючи всі можливі шляхи побудови послідовностей станів. Це запишемо наступним чином:

$$= P(Y_0, \ldots, Y_n; A, B, \pi(0)) = \sum_{j_0=1}^{J} \ldots \sum_{j_n=1}^{J} \pi_{j_0}(0) b_{j_0}(0) \prod_{l=1}^{n} a_{j_{l-1}|j_l} b_{jl}(l). \tag{22}$$

Таким чином, скінчені вимірні розподіли ймовірностей $\{Y_n\}_{n=1}^{\infty}$ є повністю визначеними нашим вибором стохастичних матриць A, B та початковим розподілом $\pi(0)$. Отже, ми можемо використовувати повний запис для цієї моделі

$$\lambda = (A, B, \pi(0)).$$

Різноманітні статті та навчальні посібники про ПММ містять одну незначну відмінність, що стосується формулювання $P(o)$ в (23). У статті Левінсона [4] та статті 1960 року Баума та Петрі була наведена наступна формула:

$$P(Y_0,\dots,Y_n,X_0,\dots,X_n;\lambda)=\pi_{j_0}(0)\cdot\prod_{l=1}^{n}b_{jl}(l)\cdot\prod_{l=1}^{n}a_{j_{l-1}|j_l}\,, \tag{23}$$

яка означає, що не існує розповсюдження в початковому стані або що початковий стан є станом «мовчання» (не будемо забувати, що в цих статтях головним застосуванням цього математичного апарату було розпізнавання голосу). З іншого боку, це говорить про те, що завжди відомо, що початковий стан буде тим же самим. Загальний вигляд цієї формули (23) веде початок з роботи Рабінера [5]. Так що потрібно ретельно розглянути та порівняти різноманітні варіанти формул для ПММ.

Джелайнек визначав приховані марківські моделі в термінах розподілу розповсюдження, яке є функцією від переходу стану. Формально це можна подати так, що $b_j(k)=P(Y_n=o_k|X_n=j)$ замінюється на $b_{i|j}(k)=P(Y_n=o_k|X_{n-1}=i,X_n=j)$. Легко побачити, що це твердження повністю тотожне тому, що використовувалось вище.

Коли ми кажемо «стохастичний механізм», це припускає наступні кроки генерації послідовностей в $\Re^{N+1}$:

1. Вибір початкового стану $X_0=j_0$, яке відповідає розподілу $\pi(0)$.
2. Встановити $n=0$.
3. Обрати $Y_n=o_k$, яке відповідає розподілу ймовірностей появи спостережень в стані $j_n-b_{j_n}(k)$.
4. Перевести до нового стану $X_{n+1}=j_{n+1}$, яке відповідає розподілу ймовірностей переходу до цього стану – $a_{j_n|j_{n+1}}$.
5. Збільшити n на одиницю та повернутися до кроку 3, якщо $n\le N$, або перервати роботу алгоритму в протилежному випадку.

Наведена вище процедура може використовуватися як генератор вибірки, так і як *модель для генерації даної послідовності* відповідною ПММ.

Тепер розглянемо *діаграми впливу* та *нестандартні ПММ*.

Стандартні приховані марківські моделі можна розглядати як граф, наведений на рис. 1. Цей вид графів відомий під назвою *діаграми впливу* або мережі довіри (див. у Сміта [6]). Кожний вузол у графі зображає випадкову величину, яка описує стан X_n або результат спостережень Y_n в деякий момент часу n. На рис. 1 ребра (стрілки) зображують напрям впливу. Граф зображує модель припущень. Умова марковості означає, що якщо нам відомо, який стан j був відвіданий у момент часу n, то ніяка інша інформація з минулого неважлива для майбутнього. На рис. 1 X_n відокремлює попередню змінну від майбутньої: після видалення X_n з графу змінні X_m ($m>n$) стануть від'єднаними від змінних X_k ($k<n$).

В термінах діаграм впливу можна легко осягнути та працювати не тільки у межах стандартних ПММ. Приклади нестандартних моделей подані на наступних рисунках: на рис. 2 зображена авторегресивна ПММ, на рис. 3 – спарена ПММ, на рис. 4 – факторна ПММ. Ключем для розуміння зображеного повинні стати круги, які мають у собі стани прихованого ланцюжку, та трикутники з виданими на відповідні моменти часу станами. Детальний розгляд факторних ПММ, поданих на рис. 4, дається у Грахрамані та Джордана [7]. Бойз розглядав у своїй роботі [8] застосування авторегресивних ПММ в аналізі біологічних послідовностей.

Для згаданої вище родини стандартних прихованих марківських моделей існує *три основні проблеми*, вирішення яких необхідно для використання ПММ для аналізу часових рядів.

1. Оцінка або проблема кількісного показника.
2. Проблема декодування.
3. Оцінка результату або проблема навчання.

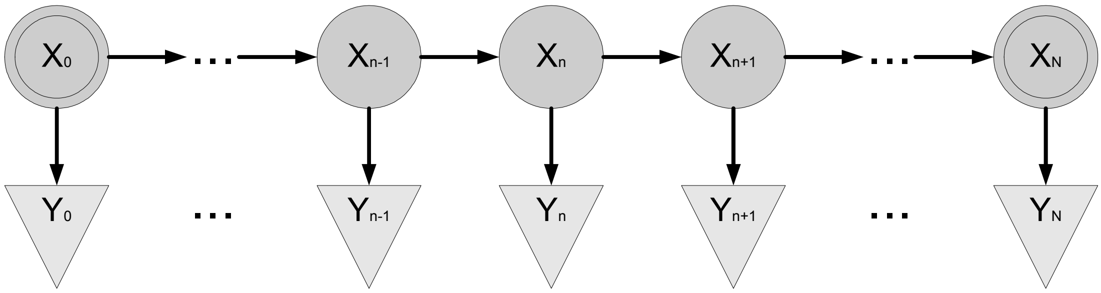


Рисунок 1 – Діаграма впливу: стандартна ПММ

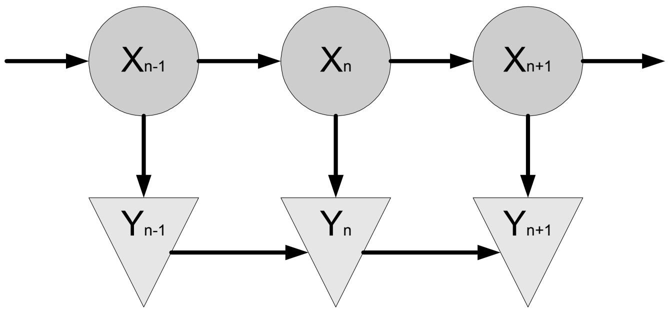


Рисунок 2 – Діаграма впливу: авторегресивна ПММ

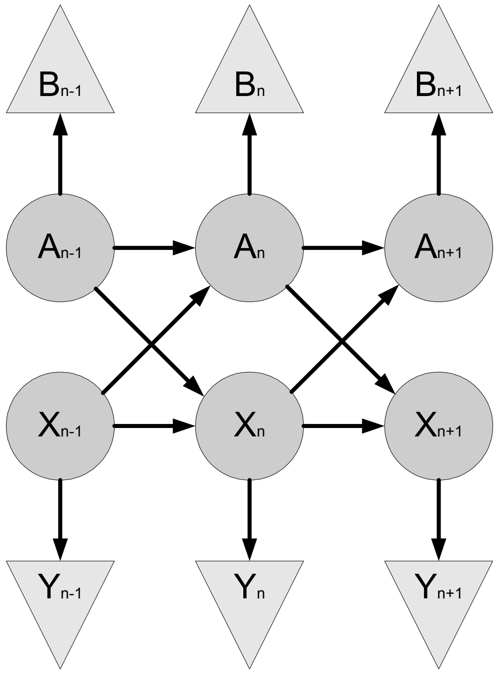


Рисунок 3 – Діаграма впливу: спарена ПММ

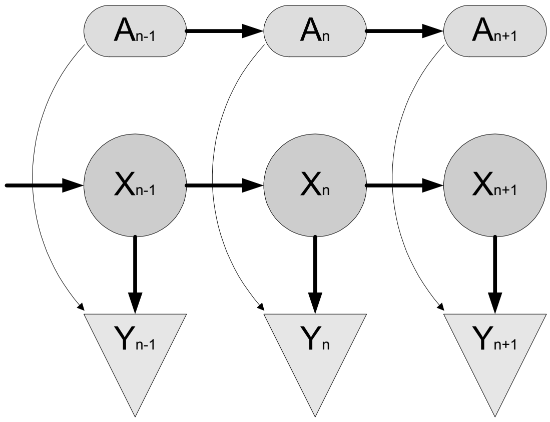


Рисунок 4 – Діаграма впливу: факторна ПММ

Оцінка результату для заданої досліджуваної послідовності $O = o_0 \ldots o_n$ має в своїй основі пошук моделі:

$$\lambda = \left(A, B, \pi(0)\right),$$

яка визначає найбільш ймовірну модель для породження заданої послідовності $o_0 \ldots o_n$, що також отримала назву тренувальної послідовності.

Для вирішення проблеми оцінки результату ми повинні піддати аналізу наступне.

1. Максимальна правдоподібність $\hat{\lambda}_{ML}$ визначається так:

$$\hat{\lambda}_{ML} = \arg\max_{\lambda} P\left(Y_0 = o_0, \ldots, Y_n = o_n; \lambda\right).$$

Ймовірність у правій частині тотожності задана в (23). Цей вираз може мати різні ступені складності в залежності від реальних виразів для умовних розподілів в B. $\hat{\lambda}_{ML}$ – для стандартної ПММ, обчисленої за допомогою *алгоритма Баума-Уелша*, який оснований на тих самих процедурах, що й алгоритм максимального очікування для скінчених сумішей.

2. Навчання максимуму апостеріорі $\hat{\lambda}_{MAP}$:

$$\hat{\lambda}_{MAP} = \arg\max_{\lambda} P\left(Y_0 = o_0, \ldots, Y_n = o_n; \lambda\right)\phi(\lambda),$$

де $\phi(\lambda)$ – щільність апріорного розподілу.

3. Мінімальна розрізнююча інформація, що була введена Ефараїмом [9], базується на мінімізації відповідної відстані Куллбека.

4. Гладкі алгоритми навчання. Гладка параметризація λ робиться для того, щоб використовувати метод градієнтного спуску. Технологія була розвинута Балді [10].

5. Навчання за Вітербі. Якщо ми встановимо, що

$$V^*(\lambda) = \max_{j_0 \cdots j_n} P\left(X_0 = j_0, \ldots, X_n = j_n; Y_0 = o_0, \ldots, Y_n = o_n; \lambda\right),$$

й спробуємо знайти таке λ, що максимізує $V^*(\lambda)$, тоді вже ми можемо говорити про навчання за Вітербі. Цей метод був розвинений та проаналізований Джуангом та Рабінером [11].

Умову для обчислювання оцінки та навчання в стандартних та нестандартних ПММ сформулював Люк [12].

Сміт та інші [13] ввели загальну структуру графічних моделей для імовірнісних незалежних мереж, завдяки яким можна вивести чисельні алгоритми для кількісної оцінки та проблеми вирівнювання в інший спосіб, ніж це було представлено в цьому тексті. Технологія імовірнісних незалежних мереж може також використовуватися при вирішенні цих проблем для ряду нестандартних ПММ.

Висновки

В цій статті були досліджені основні результати щодо використання апарату прихованих марківських моделей для аналізу часових рядів. Були розглянуті особливості використання апарату марківських ланцюжків для аналізу часових рядів. Також ми розглянули приховані марківські моделі та пов'язані з ними діаграми впливу.

В результаті проведених теоретичних та практичних експериментів було показано, що має сенс застосування ПММ для аналізу та прогнозування економічних часових рядів.

Література

1. Степанкова Г.А., Баклан І.В. Деякі підходи до навчання прихованих марківських моделей // Матеріали міжнародної наукової конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту». – Євпаторія. – 2008. – Т. 3 (ч. 2) – С. 87-91.
2. Баклан И.В., Степанкова А.А. Алгоритмы обучения скрытых марковских моделей // Системний аналіз та інформаційні технології: Матеріали X Міжнародної науково-технічної конференції (Київ, 20 – 24 травня 2008 р.). – К.: НТУУ «КПІ», 2008. – С.170.
3. Wilks S.S. (1962): Mahematical Statistics, John Wiley and Sons, New York and London.
4. Levinson S.E., Rabiner L.R., Sondhi M.M. An Introduction to the Applications of Theory of Probabilistic Functions of a Markov Chain to Automatic Speech Recognition // The Bell System Technical Journal. – 1983. – № 62. – P. 1053-1074.
5. Rabiner L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition // Proc. of the IEEE. – 1989. – № 77. – P. 257-286.
6. Smith J.Q. Influence diagramsfor statistical modelling // Annals of Statistics. – 1989. – № 17. – P. 654-672.
7. Grahramani Z. and Jordan M. Factorial hidden Markov models // Machine Learning. – 1997. – № 29. – P. 245-273.
8. Boys R.J., Henderson D. and Wilkinson D.J. Detecting homogeneous segments in DNA sequences using hidden Markov models // Applied Stistacs. – 2000. – № 49. – Part 2. – P. 269-285.
9. Ephraim Y., Dembo A. and Rabiner L.R. Minimum Discrimination Information Approach for Hidden Markov Modelling // IEEE Trans. jn Information Theory. – 1989. – № 35. – P. 1000-1013.
10. Baldi P. and Chauvin Y. Smooth On-Line Learning Algorithms for Hidden Markov Models // Neural Computation. – 2000. – № 6. – P. 307-318.
11. Juang B.H. and Rabiner L.R. The segmental K-means Algorithm for Estimating Parameters of Hidden Markov Models // IEEE Transactions on Acoustics, Speech and Signal Processing. – 1990. – № 38. – P. 1639-1641.
12. Lucke H. Which Stochastic Models Allow Baum-Welch Training? // IEEE Transactions on Signal Processing. – 1996. – № 44. – P. 2746-2756.
13. Smyth P., Heckerman D. and Jordan M.L. Probabilistic Independence Networks for hidden Markov probability models // Neural Computation. – 1997. – № 9. – P. 227-269.

И.В. Баклан, А.А. Степанкова

Вероятностные модели для анализа и прогнозирования временных рядов

Исследованы основные результаты использования скрытых марковских моделей для анализа временных рядов и связанные с ними диаграммы влияния. Рассмотрены особенности использования аппарата марковских цепочек для анализа временных рядов, а также обучение марковских цепочек.

I.V. Baklan, A.A. Stepankova

Probability Models for Analysis and Forecasting of Time Series

Researched the results of using Hidden Markov Models for time series analysis with their linkage to influence diagrams. Reviewed features of using Markovs chains for time series analysis and self-learning of Markov chains.

Стаття надійшла до редакції 21.07.2008.